

Basic Statistics

with Google Sheets

MORNING CLASS: 8.00 - 9.30 AM

October 17, 2020

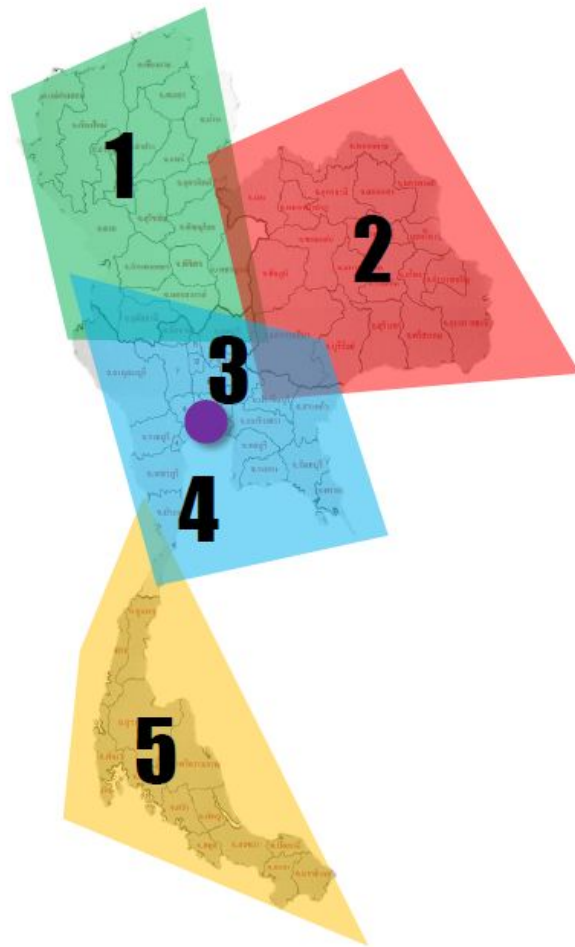
Today Agenda

1. Intro to Statistics
2. Hands-on exercises ([Google Sheets](#))
 - a. IMDb dataset
 - b. Normal distribution
 - c. Plotting distribution
 - d. Measures of central tendency
 - e. Measures of spread
 - f. Measures of position
 - g. Measures of relationship - correlation
 - h. Measures of relationship - crosstabs
 - i. Quick AB Test

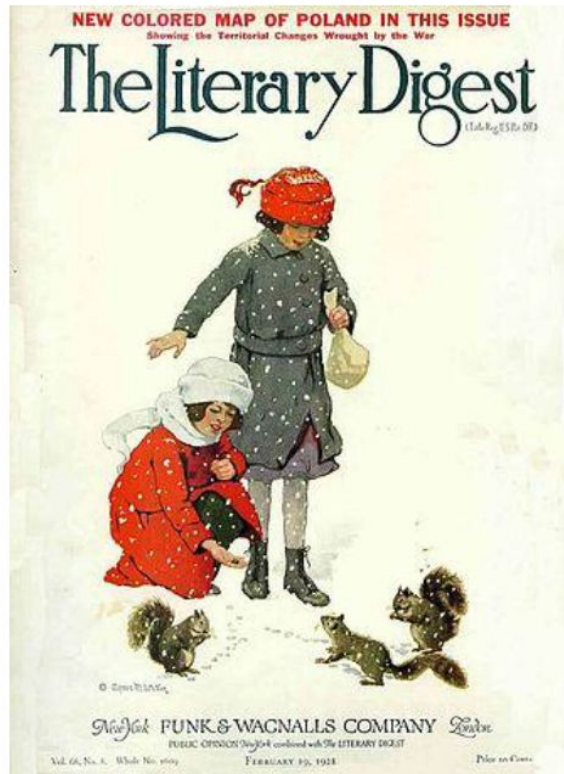


Research Design

Random / Representative



2.4 million responded



1936 predicted who was
going to win the election

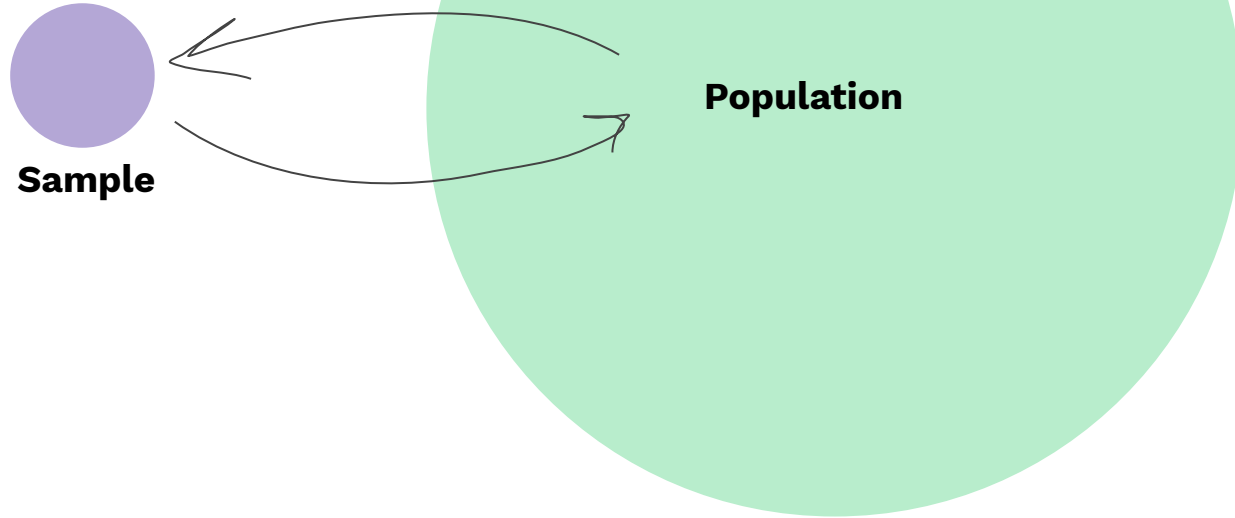
Alfred Landon **vs.** Franklin D. Roosevelt

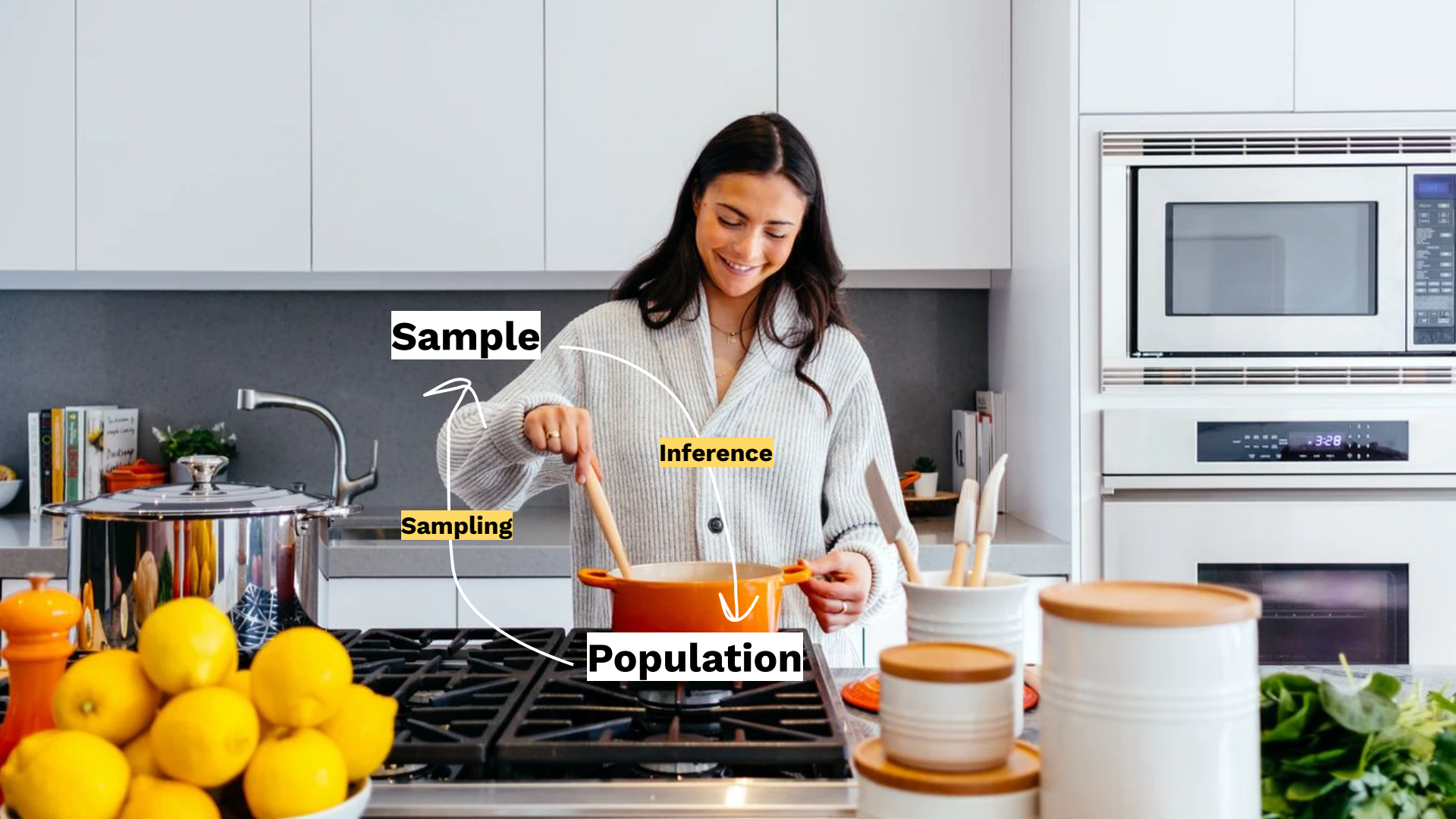
https://en.wikipedia.org/wiki/The_Literary_Digest#Presidential_poll

Quality of our sample is
of paramount importance



Inference





Sample

Inference

Sampling

Population



The image shows two women standing in front of a whiteboard. The woman on the left has long brown hair and is wearing a blue denim jacket over a colorful striped skirt. The woman on the right has long red hair and is wearing a black sleeveless top and blue jeans. The whiteboard behind them contains handwritten notes: 'DEVELOPER JOURNEY' in a circle at the top left, 'SIGN' below it, 'API LAUNCH 10' further down, 'APIs' in the center, 'PROVIDE FEEDBACK' to the right, and 'API LAUNCH 20' on the right side. A large black rectangular box with yellow diagonal lines is overlaid on the center of the image, containing the title text.

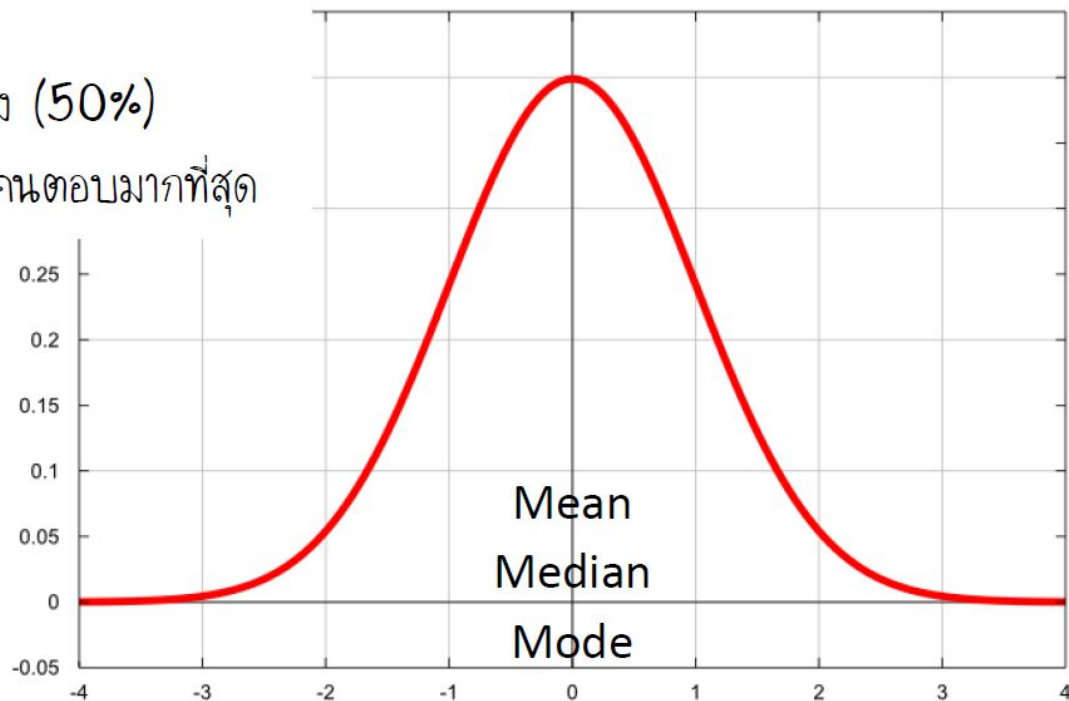
/The Descriptive Statistics

Central Tendency

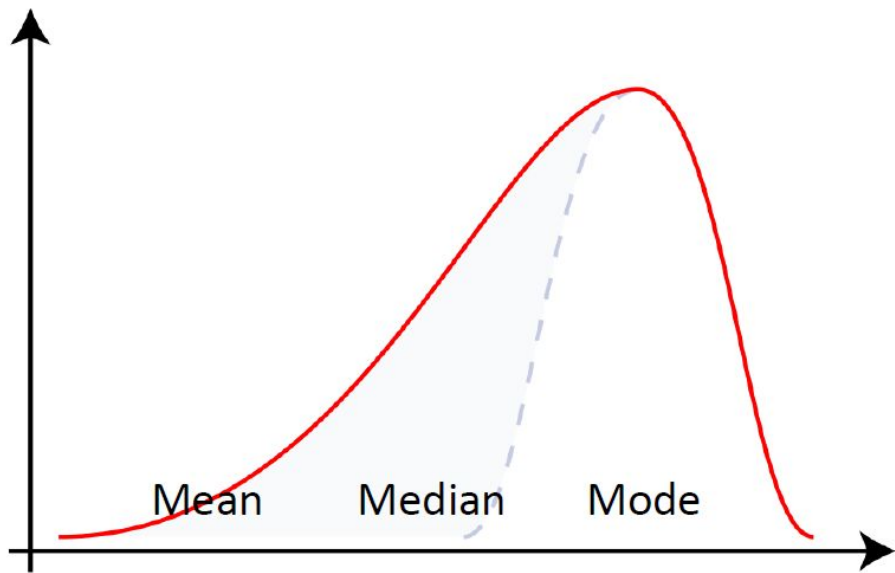
Mean = ค่าเฉลี่ย

Median = ค่ากลาง (50%)

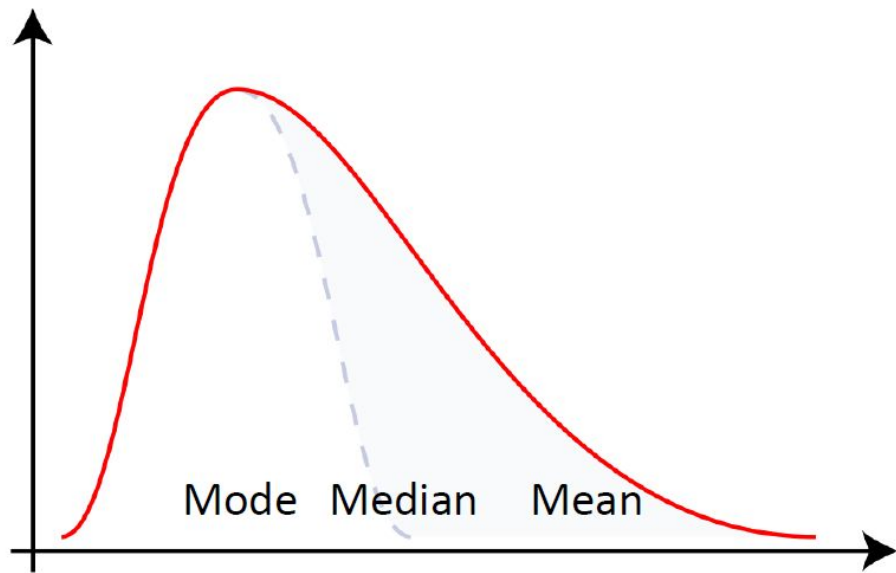
Mode = ค่าตอบที่คนตอบมากที่สุด



Skewed Distribution



Negative Skew

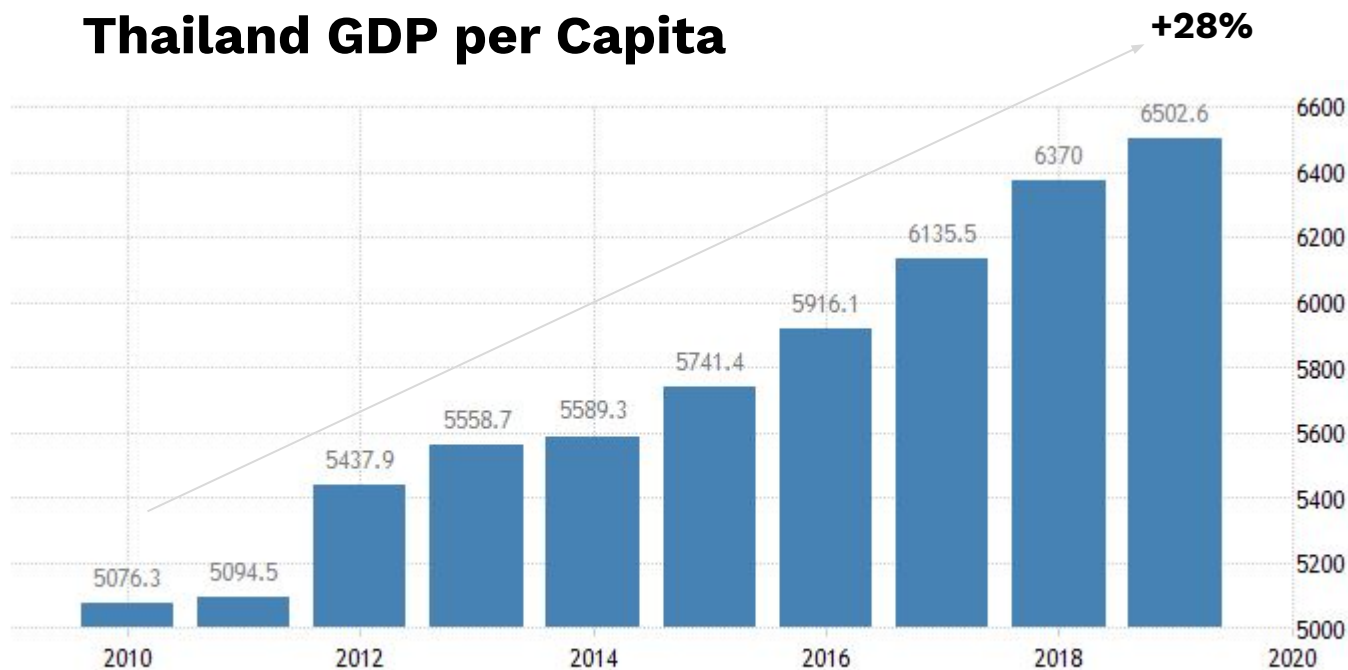


Positive Skew



When to use **mean** or **median**
depends on **the distribution**

Thailand GDP per Capita

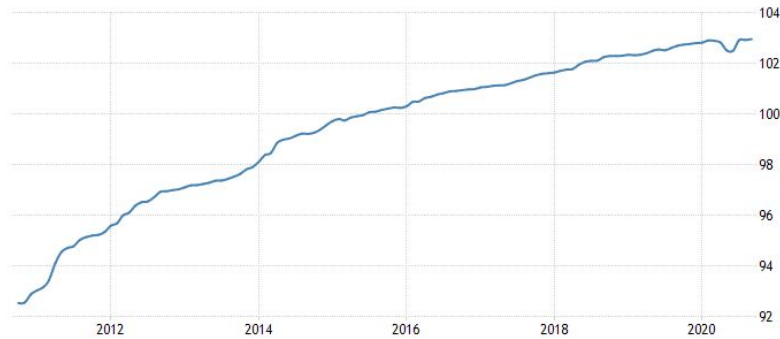


SOURCE: TRADINGECONOMICS.COM | WORLD BANK

<https://tradingeconomics.com/thailand/gdp-per-capita>

$$\mathbf{GDP} = \text{Price} \times \text{Quantity}$$

Core Consumer Prices



SOURCE: TRADINGECONOMICS.COM | BUREAU OF TRADE AND ECONOMIC INDICES, THAILAND

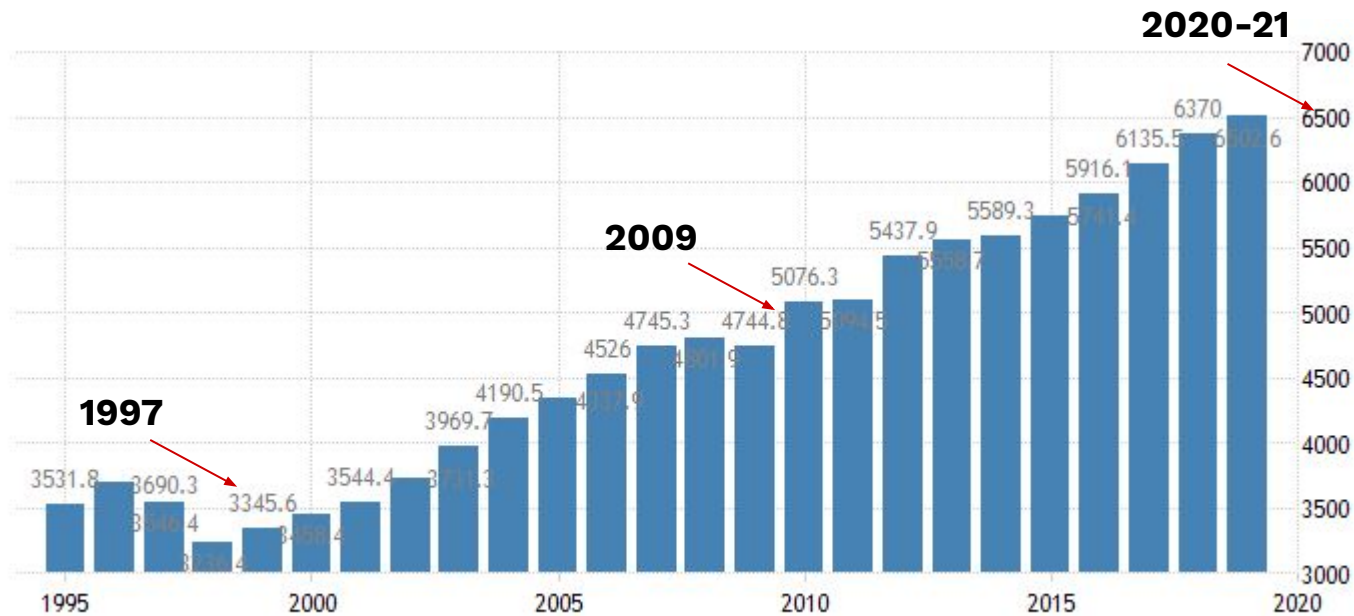
<https://tradingeconomics.com/thailand/core-consumer-prices>

Gini Index



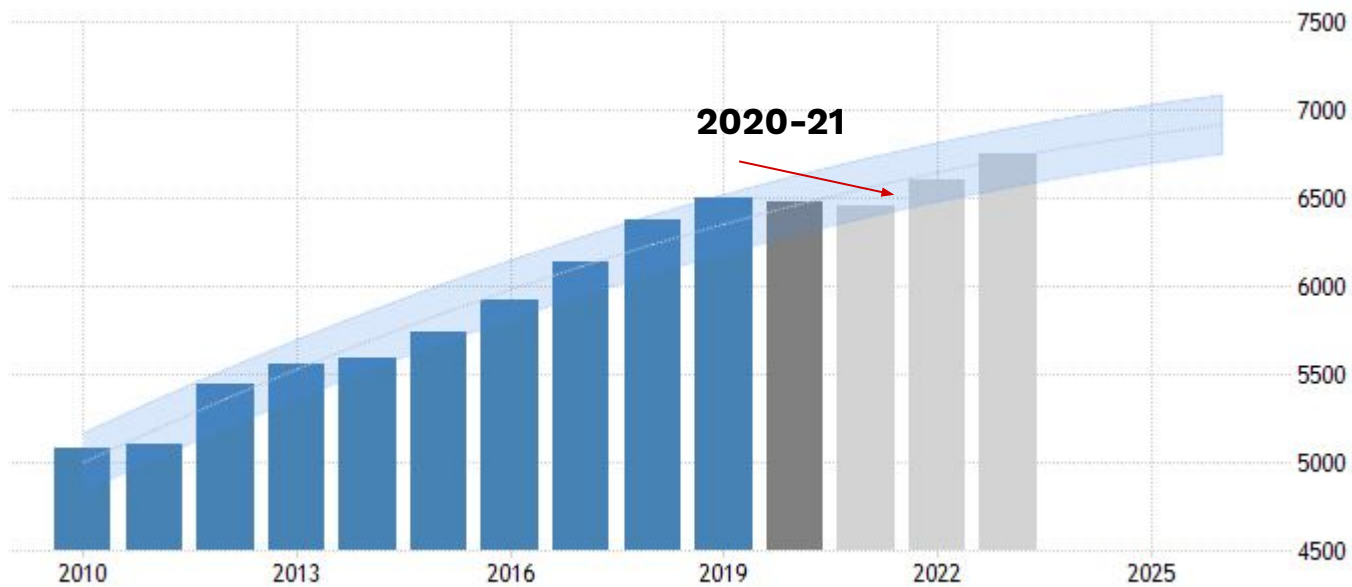
WORLD BANK | TRADINGECONOMICS.COM

<https://tradingeconomics.com/thailand/gini-index-wb-data.html>



SOURCE: TRADINGECONOMICS.COM | WORLD BANK

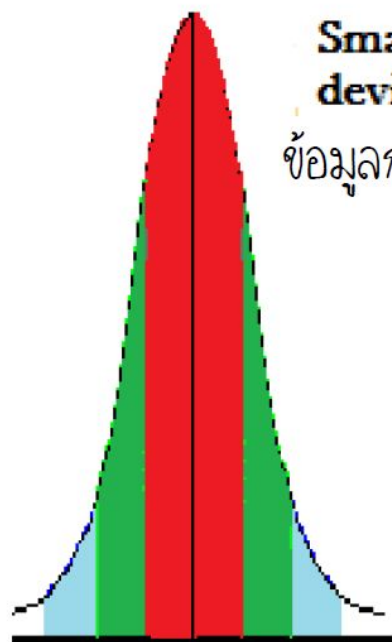
<https://tradingeconomics.com/thailand/gdp-per-capita>



SOURCE: TRADINGECONOMICS.COM | WORLD BANK

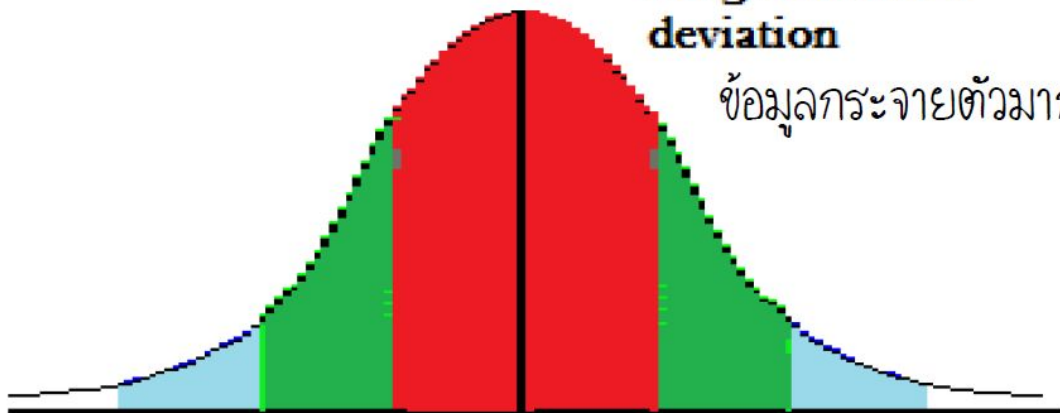
Variability

Variance and SD



**Small standard
deviation**

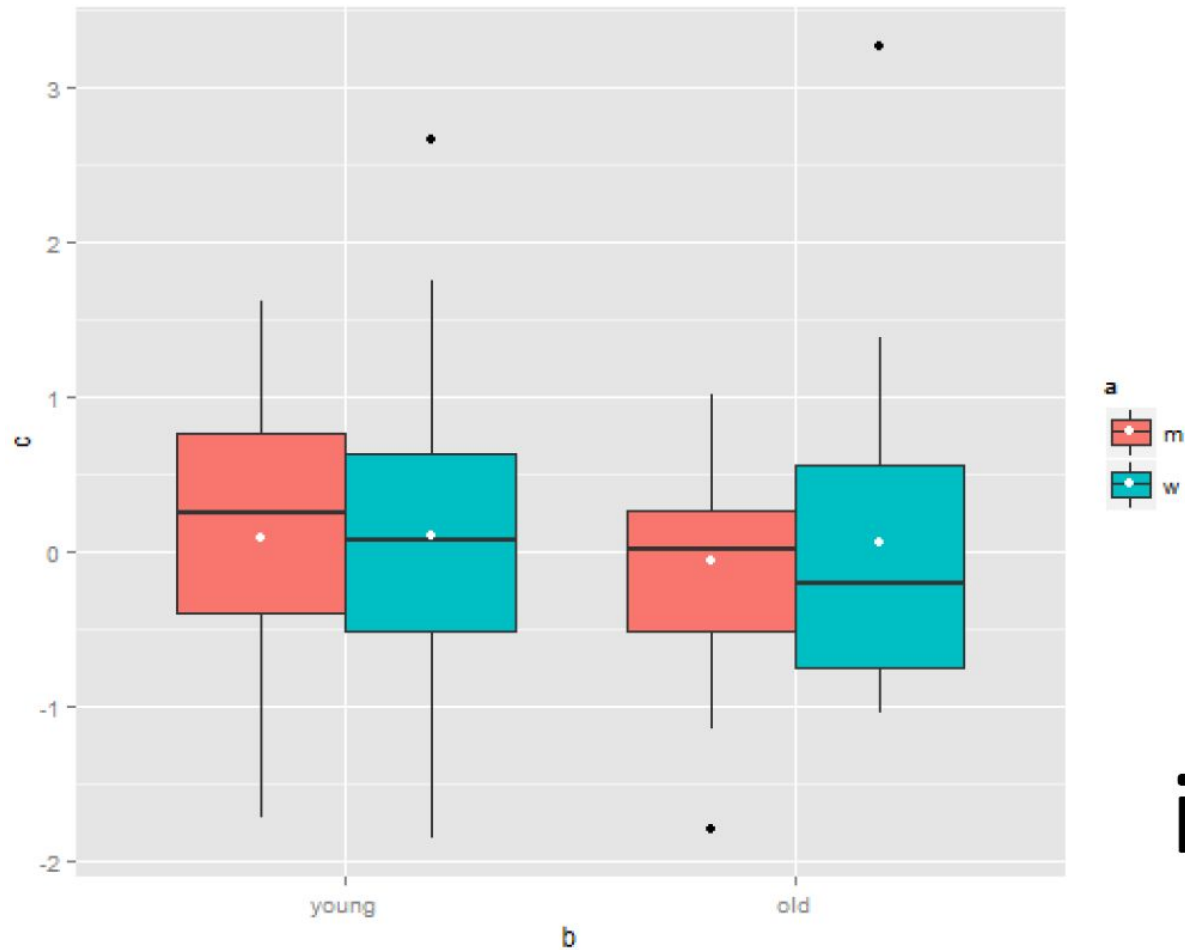
ข้อมูลกระจายตัวน้อย



**Large standard
deviation**

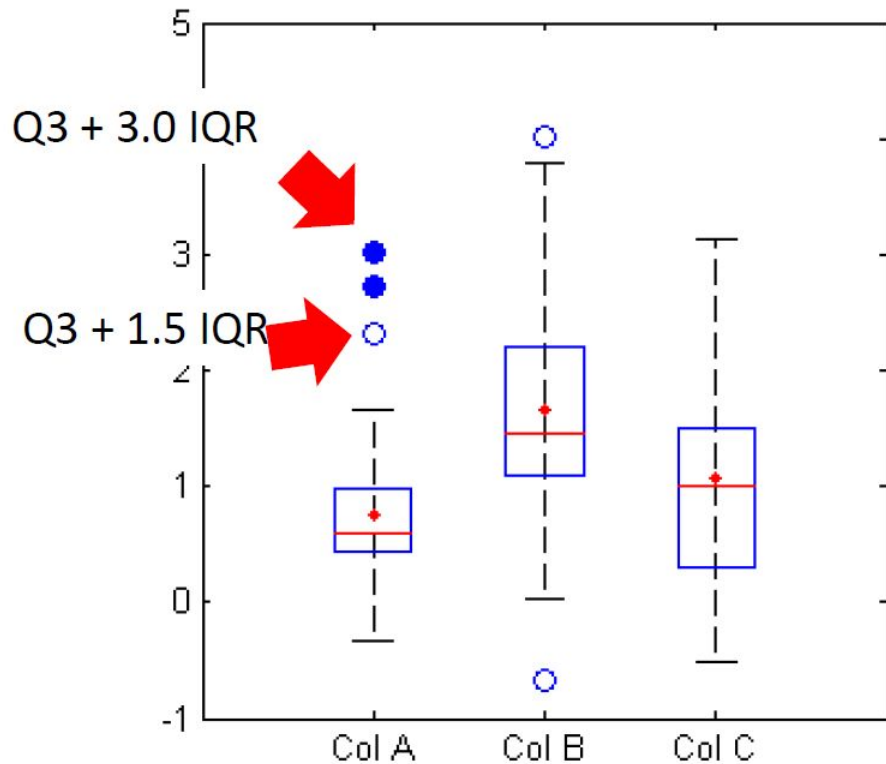
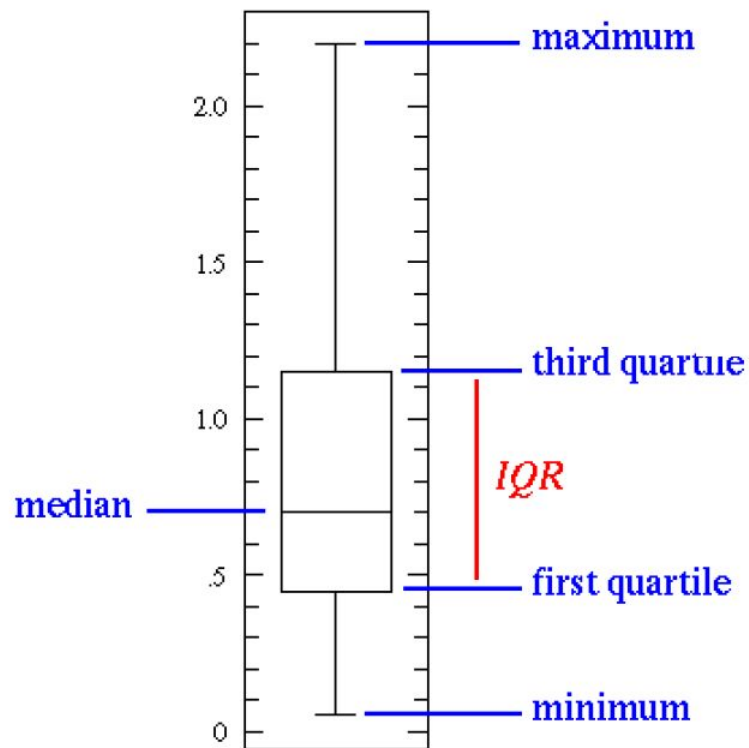
ข้อมูลกระจายตัวมาก

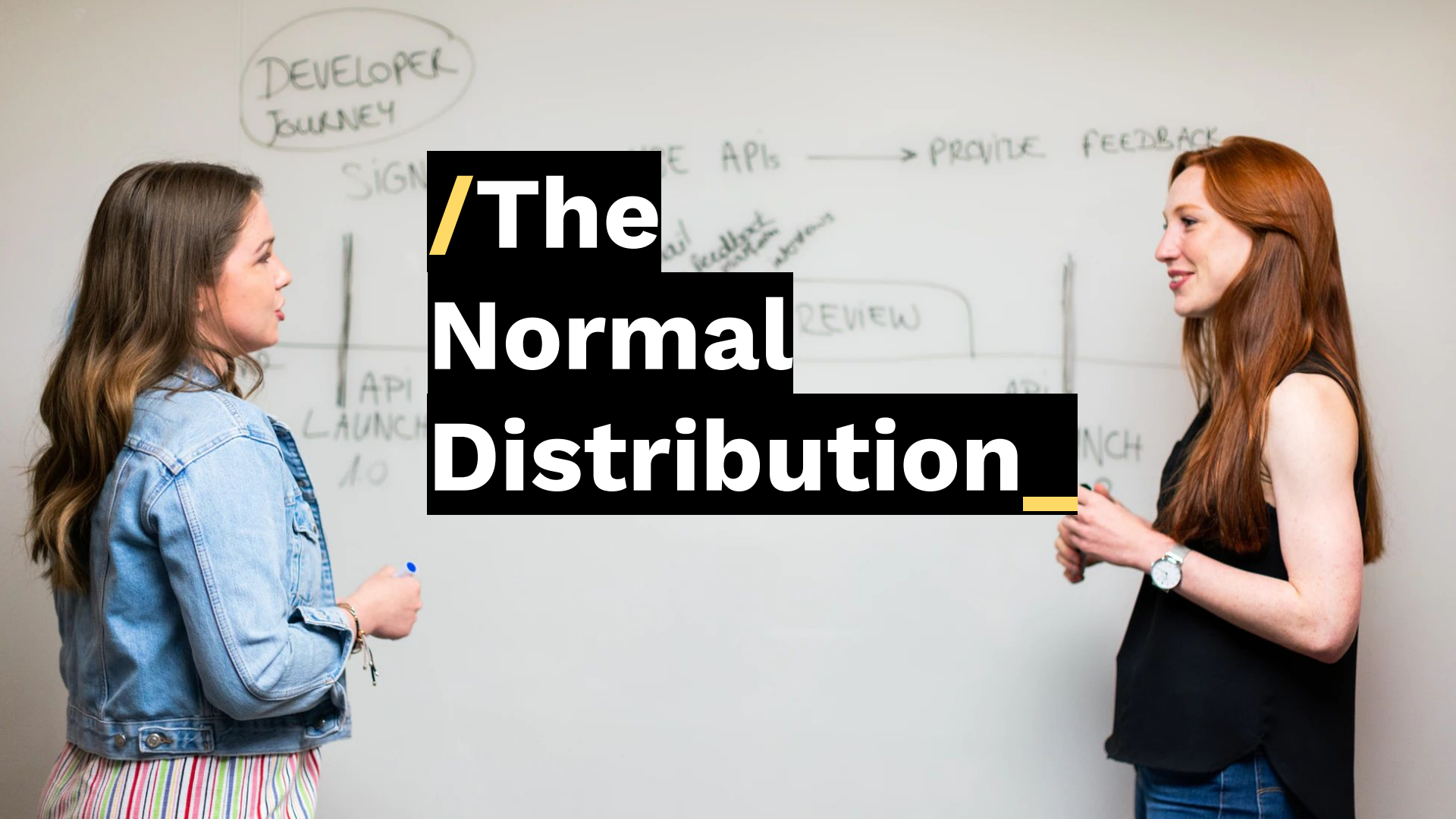
Position



Boxplot is good chart

Outliers

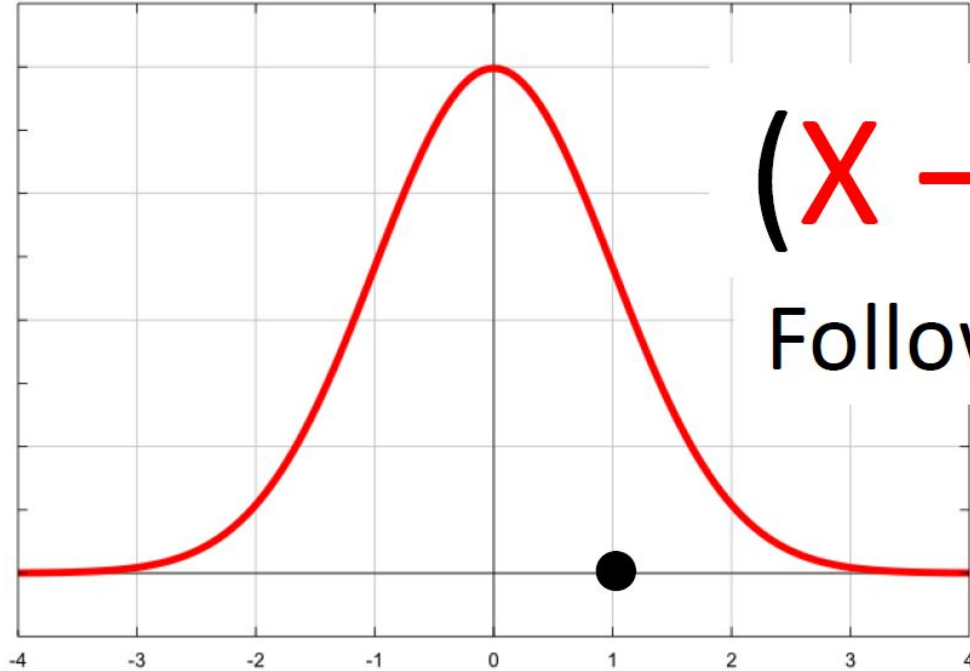




The image shows two women standing in front of a whiteboard. The woman on the left has long brown hair and is wearing a blue denim jacket over a colorful striped skirt. The woman on the right has long red hair and is wearing a black sleeveless top and blue jeans. The whiteboard behind them contains handwritten notes and diagrams. At the top left, 'DEVELOPER JOURNEY' is circled. Below it, 'SIGN' is written. To the right, 'APIs' is written with an arrow pointing to 'PROVIDE FEEDBACK'. Further down, 'REVIEW' is written in a box. At the bottom left, 'API LAUNCH' is written with '10' below it. At the bottom right, 'NCH' is partially visible. The title 'The Normal Distribution' is overlaid in the center in large white text on a black background, with yellow diagonal lines on the left and right sides of the text.

/The Normal Distribution

Z-Score



$$(X - \text{Mean}) / \text{SD}$$

Follow normal distribution

Z 90% = 1.645

Z 95% = 1.960

Z 99% = 2.575

คนที่	ส่วนสูง	Z
1	163	-0.94
2	164	-0.78
3	180	1.63
4	168	-0.18
5	174	0.72
6	173	0.57
7	161	-1.24
8	160	-1.39
9	172	0.42
10	160	-1.39
11	180	1.63
12	175	0.87
13	164	-0.78
14	164	-0.78
15	178	1.33
16	173	0.57
17	175	0.87
18	168	-0.18
19	163	-0.94
20	169	-0.03

Z can be computed by statistical software

Mean	169.20
SD	6.63

ความหมายของค่า Z คือ ความสูงของคนนี้ (178 cm) สูงกว่าค่าเฉลี่ย (mean) คนไทย อยู่ $1.33 * SD$

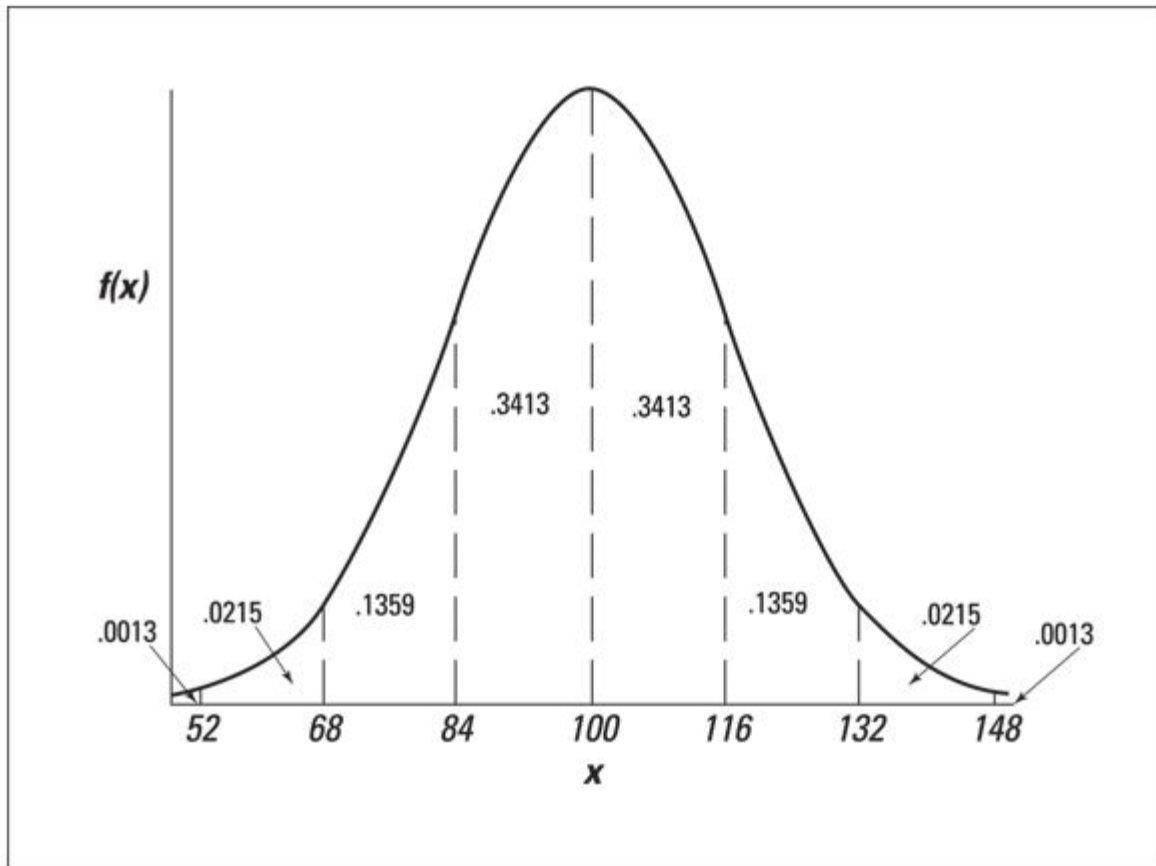
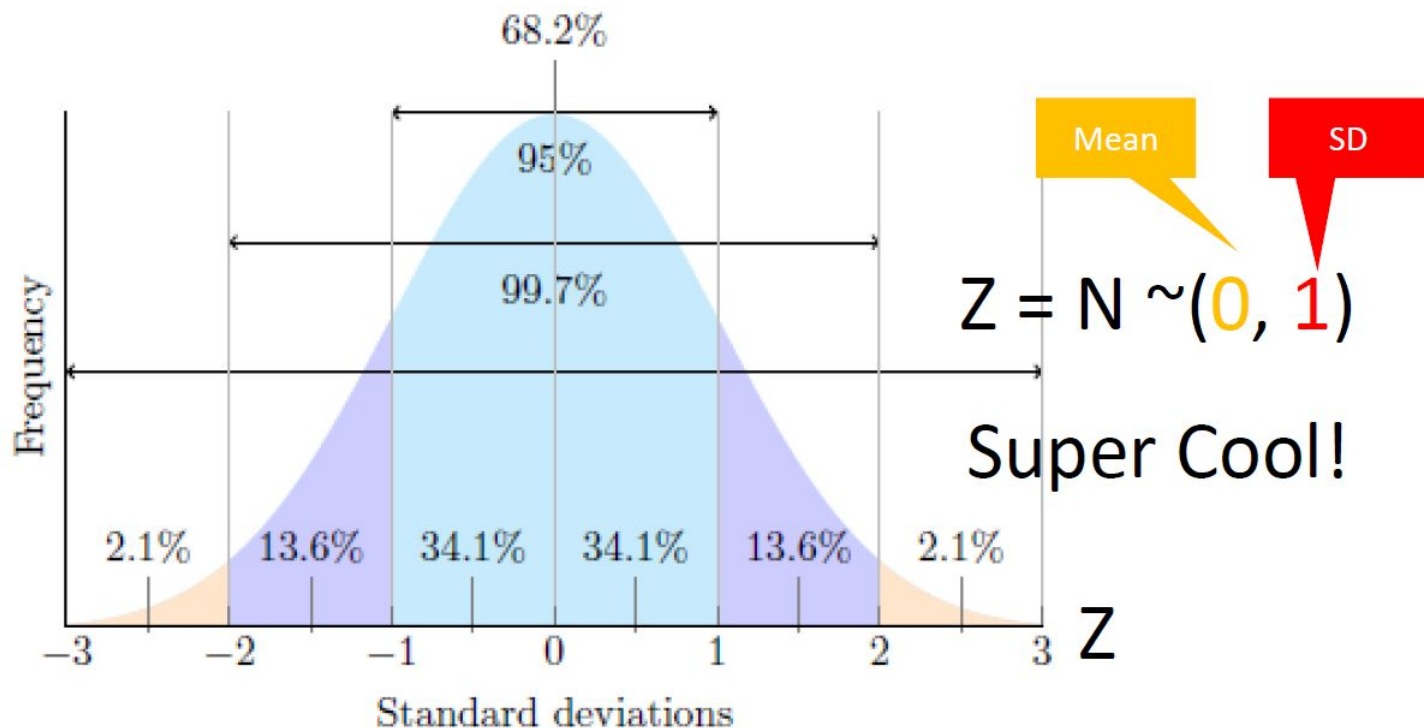


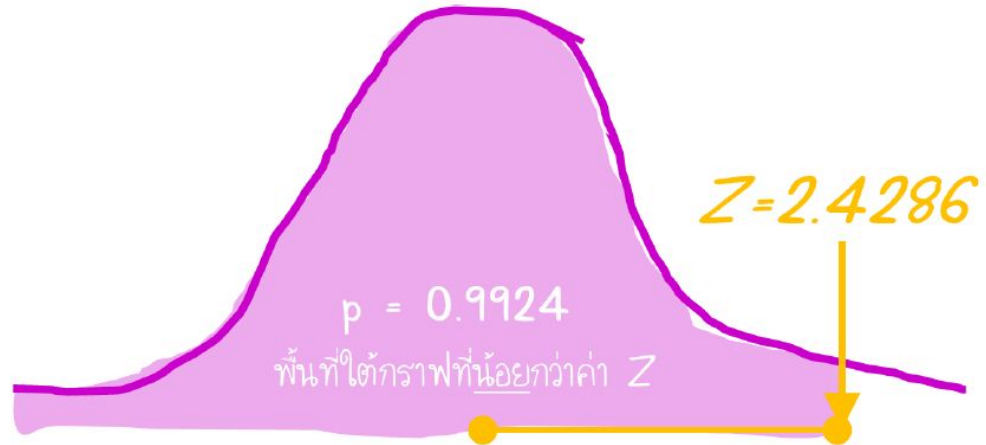
FIGURE 8-2: The normal distribution of IQ, divided into standard deviations.

Normal Distribution



Z-Score of Test score

	A	B
1	X	72
2	MEAN	55
3	SD	7
4	Cumulative	TRUE
5		
6	prob	0.9924
7	1-p	0.0076



= NORM.DIST(72,55,7,TRUE)



$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x - \mu)^2}{2\sigma^2}}$$

μ = Mean

σ = Standard Deviation

$\pi \approx 3.14159 \dots$

$e \approx 2.71828 \dots$

The diagram shows the normal distribution formula with several annotations. A red '0' is at the top right with an arrow pointing to the μ in the denominator of the exponent. A red '1' is at the bottom right with an arrow pointing to the σ^2 in the denominator of the exponent. Another red '1' is on the left with an arrow pointing to the σ in the denominator of the coefficient. The variables x and μ in the numerator of the exponent are crossed out with red lines, and a red 'Z' is placed above the $(\quad)^2$ term.


$$y = \frac{1}{\cancel{\sigma} \sqrt{2\pi}} e^{-\frac{(\cancel{x} - \cancel{\mu})^2}{2\cancel{\sigma}^2}}$$

0

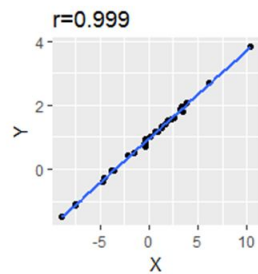
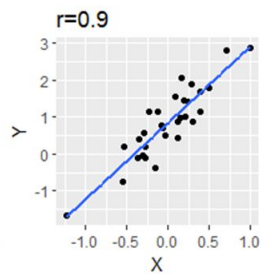
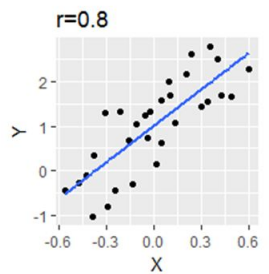
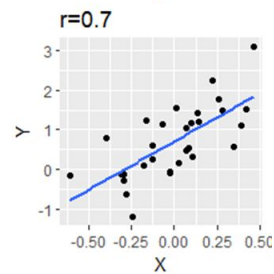
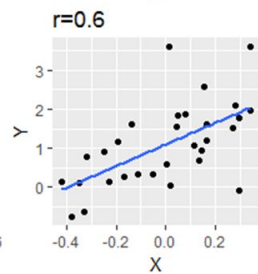
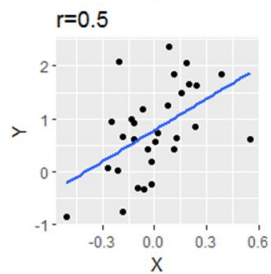
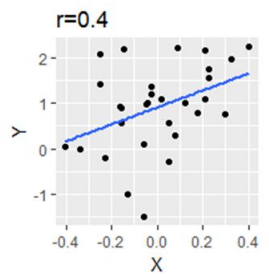
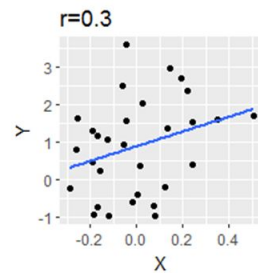
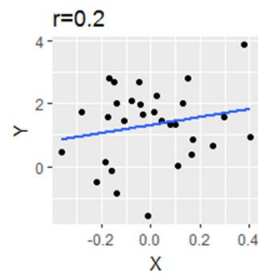
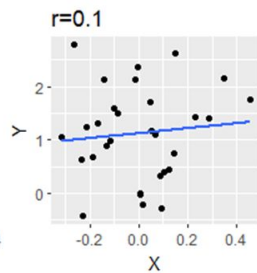
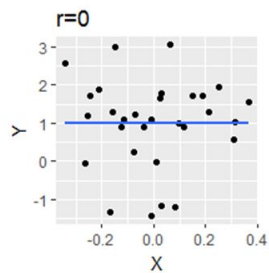
1

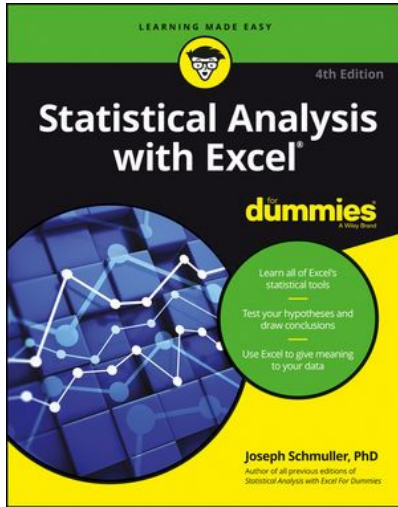
1

$\mu = \text{Mean}$
 $\sigma = \text{Standard Deviation}$
 $\pi \approx 3.14159 \dots$
 $e \approx 2.71828 \dots$

A photograph of two women standing in front of a whiteboard. The woman on the left has long brown hair and is wearing a blue denim jacket over a colorful striped skirt. The woman on the right has long red hair and is wearing a black sleeveless top and blue jeans. The whiteboard behind them has handwritten notes, including 'DEVELOPER JOURNEY' circled at the top left, and a flow diagram with 'USE APIs' and 'PROVIDE FEEDBACK' connected by an arrow. Other faint notes include 'email feedback system' and 'LAUNCH'. A large black text box with white text is overlaid in the center, reading '/The Measures of Relationship'.

/The Measures of Relationship





Book Reference

Statistical Analysis with
Excel for dummies (2016)

Basic Statistics

with Google Sheets

MORNING CLASS: 8.00 - 9.30 AM

SCHOOL OF FULL STACK
[HTTPS://DATAROCKIE.COM](https://datarockie.com)

October 17, 2020